DOCUMENT RESUME

ED 480 074 CG 032 647

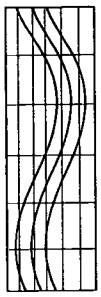
AUTHOR	Erford, Bradley T.; Moore-Thomas, Cheryl
TITLE	Testing FAQ: How to Answer Questions Parents Frequently Ask about Testing.
PUB DATE	2003-08-00
NOTE	23p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Accountability; *Educational Testing; *Parent Participation; Parent School Relationship; *Test Wiseness; *Testing

ABSTRACT

Educational accountability demands that students take tests. Parents and guardians, being committed to their children's academic success, often ask teachers and other educators questions about tests and testing procedures. This chapter provides practical, straightforward responses to many of the questions parents and guardians ask about testing. (Contains 19 references and 2 tables.) (Author)

Testing FAQ: How to Answer Questions Parents Frequently Ask About Testing

By
Bradley T. Erford
Cheryl Moore-Thomas



Chapter 39
Testing FAQ
How to Answer Questions Parents Frequently Ask About Testing

Bradley T. Erford & Cheryl Moore-Thomas

Educational accountability demands that students take tests. Parents and guardians, being committed to their children's academic success, often ask teachers and other educators questions about tests and testing procedures. This chapter provides practical, straightforward responses to many of the questions parents and guardians ask about testing.

The Purpose of Testing

What is the purpose of the tests my child is taking? Can tests determine how well my child is doing in school?

All of us have taken tests. More than ever, school-age children are being required to take many different kinds of tests. Much of the testing in today's schools may be attributed to national concern for accountability in public education resulting from Goals 2000: Educate America Act. Goals 2000 provides a framework for educational reform by improving the quality of learning and teaching in the classroom, and assisting in the development of quality assessment measures. Testing is essential to the very purpose of education (Coffman & Lindquist, 1980).

In general, the main purpose of testing is to benefit students. Tests help educators and parents identify student strengths and areas needing improvement. Educators can use information from tests to plan lessons and design curriculum that meet the needs of all students. Tests can also help evaluate and improve schools or entire school systems. Thus, testing information is crucial for educational accountability (Educational Testing Service, 1999; Eissenberg & Rudner, 1988).

Classroom tests are probably the most common type of tests students take. These tests are often teacher-made and cover a specific body of knowledge. Classroom tests may be short and clear-cut, like

weekly spelling tests, or they may be fairly involved, like unit tests in social studies or science or even high school final exams. Classroom tests are given to help educators, parents, and guardians assess what students have learned.

Students may also take standardized tests. Standardized tests are used to help measure student ability or achievement. Standardized ability tests measure students' capacity to learn, whereas standardized achievement tests measure what students have learned about a particular subject. Classroom teachers do not create standardized tests. Commercial test publishers develop most of these tests, which are administered in the same way for all test takers. This standardization is what makes these tests a powerful tool in assessment. Standardization enables comparisons to be made among individuals and schools.

There are two basic kinds of standardized tests: norm-referenced and criterion-referenced. *Norm-referenced* tests compare students' performance to that of their peers, while *criterion-referenced* tests compare or measure students' performance against particular standards. On norm-referenced tests, students' scores are compared to the scores of the original group of students who took the test, called a *norm group.* Norm-referenced tests may answer questions such as, "How does my child's understanding of word meanings compare to that of her peers?"

Student performance on criterion-referenced tests is measured against a specific set of skills or objectives or against an established criterion for passing or mastery. Criterion-referenced tests may answer the questions, "Does my child know the meaning of the word 'periodic'?" or "Does my child know how to add two-digit numbers with regrouping?"

Testing is an important part of the education process. Used appropriately, tests can help educators, parents, guardians, students, and other stakeholders make critical decisions about educational programming and services. Tests alone, however, do not give the complete picture of any student's knowledge or ability. They give a single snapshot of student performance, a single piece of the assessment process (Bagin, 1989; Coffman & Lindquist, 1980; McMillan, 2000; Salvia & Ysseldyke, 2001).

The Content of Tests

Who decides what questions go on the tests? Shouldn't the teacher have a role in question selection?

Test authors and publishers decide test questions; however, test authors rely on content specialists who review applicable national, state, and local standards and curricula (including textbooks) to determine what comprises the domain of knowledge to be tested and to select information that is important for students to know. This ensures a test has *content validity* (Salvia & Ysseldyke, 2001). Many content specialists are current or former teachers.

On standardized tests designed in cooperation with local school systems or state departments of education, selected classroom teachers often have a role in selecting learning objectives that guide question selection. Teachers even submit questions for consideration. Thus, although teachers may not select the actual questions, they often help prioritize the content that guides question selection. In this way, content specialists and teachers work together to help determine what content is assessed, but teachers do not know the specific questions that appear on a test, which may give their students an unfair advantage should teachers "teach to the test" (Anastasi & Urbina, 1997).

Were all the questions on the test covered in class or in the textbook? How can teachers know what to cover to prepare students for the test without teaching to the test?

Curriculum standards provide learning outcomes and objectives that guide classroom instruction. Test content is also guided by these learning objectives, which are operationalized through the test questions (Popham, 2000). School systems should choose standardized tests that have substantial overlap between the test content and school curriculum. If a school system chooses a test that has only a 75 percent overlap between test content and learning objectives, their students will fare worse than students in a school system with a 100 percent overlap not because the former school system has inferior teachers or students, but because about 25 percent of what the test measures is *not taught.* When a curriculum and the test are in total alignment, the burden falls on the teacher to cover all curricular content in an efficient manner. Failure to do so will lead to lower student performance.

Teaching to the test is a problem only if the teacher has advance warning of specific questions that will appear on a test. If a teacher knows that certain content is always emphasized on a particular test, it is appropriate to emphasize that content area in instruction. Likewise, if certain content is regularly de-emphasized on a test, less attention to that content in the classroom may be warranted. It is incumbent on the school system and test publisher to ensure that standards, curriculum, and assessment are well aligned and that all objectives are assessed in the correct proportion. This way appropriate textbooks and classroom activities can be determined.

The Protection of Test Content

Why can't I get a copy of standardized test questions to help my child study ahead of time, like we do for spelling or math?

Test content is protected for a variety of reasons (Anastasi & Urbina, 1997; Cohen & Swerdlik, 1999; Kaplan & Saccuzzo, 2001; Salvia & Ysseldyke, 2001; Thorndike, 1997). Most standardized tests must be administered, scored, and interpreted by individuals with specialized education, training, and experience. Among other things, these individuals must be able to select an appropriate test, administer and score the test accurately, and interpret the score. Test content must be protected because the results will not yield a valid estimate of current abilities if the person taking the test knows the questions and answers beforehand. Standardized tests differ from classroom spelling or math tests in this regard because the content domain on a spelling or math test is usually revealed and studied in close proximity to the test. Studying for a classroom test is generally easy and the test result is compared to a grading criterion (i.e., A, B, C, and so on). On a norm-referenced test, the content is revealed over a period of several semesters or years and preparing for it is therefore much more difficult. In addition, the student's score is compared with those of the same age or grade rather than with a grading criterion.

Finally, results of standardized tests are usually less obvious or understandable than those of teacher-made tests; effectively communicating the results to parents and teachers requires specialized training. Effective communication of results and what to expect during testing helps dispel anxiety, maximize performance, and familiarize the student with the testing procedures.

Preparing for Tests

How can I help my child prepare for tests?

Tests can cause anxiety in students. This anxiety can be diminished if parents, guardians, and educators help prepare students for tests. Most importantly, parents, guardians, and educators should let students know that test taking is a normal part of the educational process. Whether students are preparing to take classroom tests or standardized tests, students should view tests as an important but regular school activity. Students are also well served by knowing what material is being covered on a test and why it is important. Teachers usually help students by providing review materials or study suggestions for classroom tests. Coaching materials are available for some standardized tests. Although coaching materials such as test preparation courses may improve students' test scores, these methods often do not appreciably improve students' mastery of the domain of information being assessed.

Students' performance may also be enhanced if they are familiar with test-taking procedures (Educational Testing Service, 1999). Test-taking procedures include understanding test response format (e.g., multiple choice, essay, true-false), test length, and test directions. Although it is appropriate for students to be familiar with test-taking procedures, it is not appropriate for them to prepare for tests by practicing with the actual test or practicing on a published parallel form of the test (Mehrens, 1989). Students should also be aware of factors that may affect their scores. For example, some tests penalize students for guessing or not answering all questions. Other tests require that students demonstrate their preliminary calculations or show their work in other ways to earn top scores.

Perhaps the best way to prepare students for tests is to consistently monitor their progress, assist them in developing strong study habits, and ensure they approach each testing situation well rested and well fed (Bond, 1996).

The Meaning of Scores

What do all the scores on my child's testing report mean? What are percentile ranks, stanines, and grade equivalents?

Simply put, norm-referenced, standardized scores are all based on the properties of a normal (bell-shaped) curve. In this way, there is

consistency in the transformation of scores as long as distributions are normal and standardization samples are similar in constitution (Cohen & Swerdlik, 1999; Thorndike, 1997). Figure 1 illustrates this similarity in score transformation. Notice that a deviation IQ of 100 (M = 100, SD = 15) will always be equivalent to a percentile rank of 50, T-score of 50, scaled score of 10, and stanine of 5. Likewise, a deviation IQ of 130 will always be equivalent to a percentile rank of 98, T-score of 70, scaled score of 16, and stanine of 9. Table 1 provides many transformations for the standardized scores commonly used in education.

Figure 1. Score Transformation Based on Normal Curve

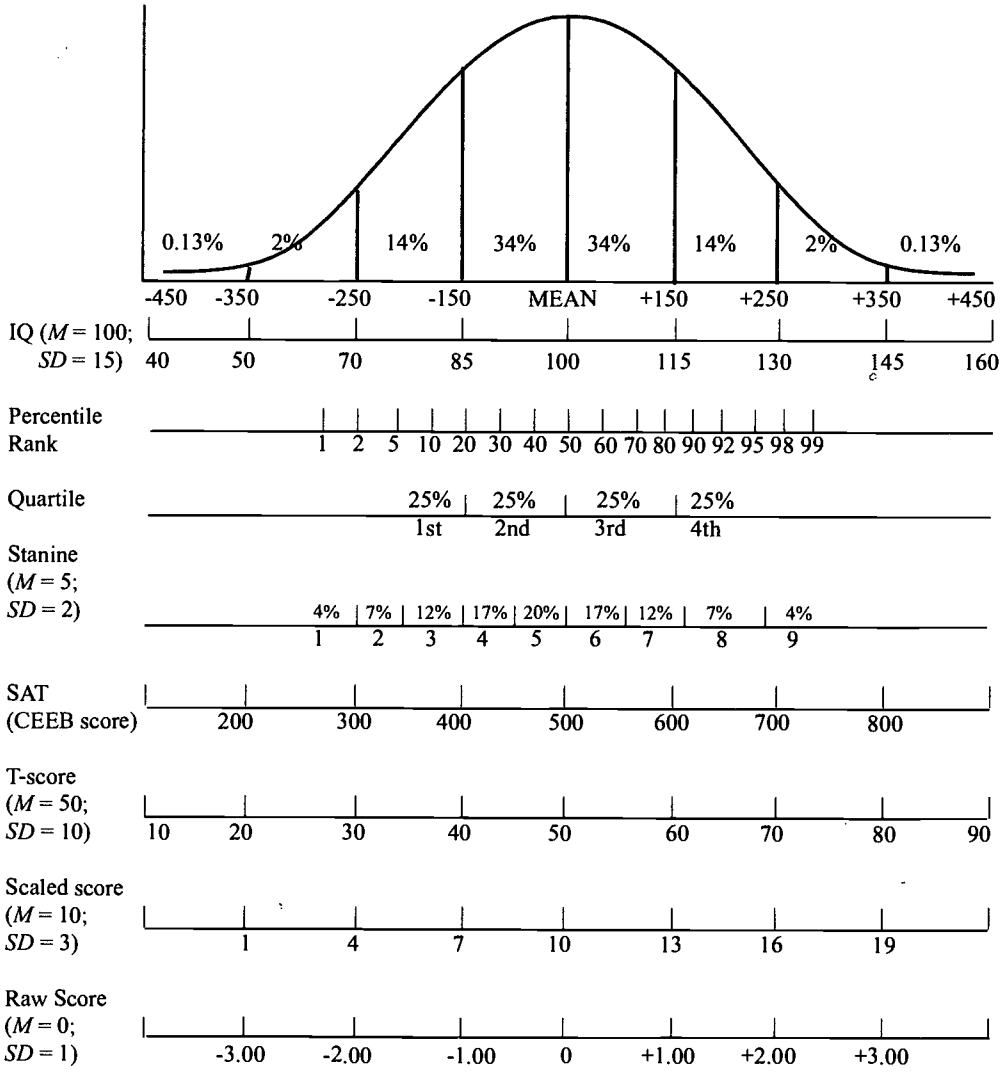


Table 1. Correspondences Among Deviation IQ, Stanine, Percentile Rank, Scaled Score, and Interpretive Range

	Deviation IQ	Stanine	Percentile Rank	Scaled Score	Interpretive Range
Mean	100	5	---	10	—
Standard	15	2	—	3	—
	55	1	<1	1	MD
	56	1	<1	1	MD
	57	1	<1	1	MD
	58	1	<1	2	MD
	59	1	<1	2	MD
	60	1	<1	2	MD
	61	1	1	2	MD
	62	1	1	2	MD
	63	1	1	3	MD
	64	1	1	3	MD
	65	1	1	3	MD
	66	1	1	3	MD
	67	1	1	3	MD
	68	*1*	*2*	*4*	*MD*
	69	*1*	*2*	*4*	*MD*
	70	*1*	*2*	*4*	*B*
	71	*1*	*3*	*4*	*B*
	72	*1*	*3*	*4*	*B*
	73	*2*	*4*	*5*	*B*
	74	*2*	*4*	*5*	*B*
	75	*2*	*5*	*5*	*B*
	76	*2*	*5*	*5*	*B*
	77	*2*	*6*	*5*	*B*

	78	*2*	*7*	*6*	*B*
	79	*2*	*8*	*6*	*B*
	80	*2*	*9*	*6*	*LA*
	81	2	10	6	LA
	82	3	12	6	LA
	83	3	13	7	LA
	84	3	14	7	LA
	85	3	16	7	LA
	86	3	18	7	LA
	87	3	19	7	LA
	88	3	21	8	LA
	89	4	23	8	LA
	90	4	25	8	A
	91	4	27	8	A
	92	4	30	8	A
	93	4	32	9	A
	94	4	34	9	A
	95	5	37	9	A
	96	5	40	9	A
	97	5	42	9	A
	98	5	45	10	A
	99	5	48	10	A
	100	5	50	10	A
	101	5	53	10	A
	102	5	55	10	A
	103	5	58	11	A
	104	6	61	11	A
	105	6	63	11	A
	106	6	66	11	A
	107	6	68	11	A

	108	6	70	12	A
	109	6	73	12	A
	110	6	75	12	HA
	111	7	77	12	HA
	112	7	79	12	HA
	113	7	81	13	HA
	114	7	83	13	HA
	115	7	84	13	HA
	116	7	86	13	HA
	117	7	87	13	HA
	118	8	89	14	HA
	119	8	90	14	HA
	120	8	91	14	S
	121	8	92	14	S
	122	8	93	14	S
	123	8	94	15	S
	124	8	95	15	S
	125	8	95	15	S
	126	9	96	15	S
	127	9	96	15	S
	128	9	97	16	S
	129	9	97	16	S
	130	9	98	16	VS
	131	9	98	16	VS
	132	9	98	16	VS
	133	9	99	17	VS
	134	9	99	17	VS
	135	9	99	17	VS
	136	9	99	17	VS
	137	9	99	17	VS

	138	9	99	18	VS
	139	9	99+	18	VS
	140	9	99+	18	VS
	141	9	99+	18	VS
	142	9	99+	18	VS
	143	9	99+	19	VS
	144	9	99+	19	VS

MD = Mildly Deficient B = Borderline LA = Low Average A= Average
HA= High Average S = Superior VS = Very Superior

It is essential to understand the scores commonly used when reporting student standardized test scores. *Percentile ranks,* which are probably the easiest score for parents and teachers to understand, are simply an indication of where a student's performance falls compared with other students of the same age or grade comprising the norm group. To explain percentile ranks, it is helpful to visualize a lineup of 100 students of the same age or grade, with the first student in the line being the least proficient and the 100th student being the most proficient. If a student scored at the 72nd percentile rank, her score would be interpreted as follows: "Susan's math calculation score exceeded the performance of 72 percent of other students in her grade (or of her age)."

Stanines (short for "standard nines") are standard score ranges dividing the distribution into nine parts. Table 1 provides stanine equivalents associated with percentile ranks. For Susan's percentile rank of 72, her math calculation performance would have fallen into the 6th stanine. It is difficult to describe to parents how stanines are derived, and therefore stanines should be used sparingly.

Grade equivalents (GEs) are computed by determining the average raw scores obtained by students in each grade at different times during the year. Erford, Vitali, Haas, and Boykin (1995, pp. 28–29) summarize the use of grade equivalents as follows:

> Despite their popular appeal, GEs are frequently misinterpreted and most often not helpful in getting teachers and parents to understand the child's performance. This is true for several reasons. First, if a child in grade 2.0 obtains a math GE of 4.0, this does not mean he/she should be immediately placed in the fourth grade curriculum. His/

> her GE does indicate that he/she will probably be a good math student in his/her second grade class . . . Second, curriculums vary in degrees of acceleration provided. . . . Some third grade curriculums are dealing with second grade concepts, while others are accelerated to the point that fourth and fifth grade content is covered to a substantial degree . . . Finally, . . . GEs should not be viewed as a performance criterion . . . a GE of 3.9 is commensurate with the average performance of an ending third grader. Thus, it would be an unrealistic expectation for all students to achieve a GE of 3.9 or higher at the conclusion of the third grade year.

An *interpretive range* is an easily understood verbal descriptor of a student's performance. Table 2 shows interpretive ranges for comparable standard scores and percentile ranks. Using the previous example of Susan, her math calculation percentile rank of 72 falls in the Average range.

Table 2. Equivalence of Standard Scores (M = 100; SD = 15), Percentile Ranks, and Interpretive Ranges

Standard Score	*Percentile Rank*	*Interpretive Range*
130+	98+	Very Superior (VS)
120–129	90–97	Superior (S)
110–119	75–89	High Average (HA)
90–109	25–73	Average (A)
80–89	10–23	Low Average (LA)
70–79	2–8	Borderline (B)
55–69	< 2	Mildly Deficient (MD)

How are tests scored? Is it possible to score essay exams accurately?

Many objective standardized tests in mass testing programs are scored using high-speed scanners and computer programs. If the students fill out the forms correctly and the publisher's answers are keyed correctly, this scoring system is virtually error-free because objective questions (i.e., multiple choice, true-false, coded) maximize interscorer reliability. *Interscorer reliability* refers to the consistency of agreement among multiple scorers of the same set of scores. Because no scorer judgment is required on multiple-choice questions, consistency is nearly always 100 percent, except in instances of miskeyed responses or errors due to inattention. The advantage of computer scoring is that errors

due to inattention are eliminated (Salvia & Ysseldyke, 2001).

On most tests with objective-type questions, only one answer for each item is correct and the number of correct items for a student on a given subtest is summed to give a raw score. This raw score is then converted, using the appropriate norm for the child's age or grade, to a standardized score and percentile rank to indicate the student's performance in relation to his or her peers.

Essay or constructed-response exams are somewhat more complicated, as interscorer reliability can become more of a factor. Under most circumstances, a scoring rubric must be constructed and sample responses, or exemplars, developed (Popham, 2000). Most constructed-response tests must be hand-scored by a qualified examiner, and in many instances, more than one examiner. Having two or more examiners score the response provides an extra check, which boosts confidence that the score has been consistently derived. Still, interscorer reliability for constructed-response tests necessarily introduces unwanted error—usually between 5 and 20 percent, opposed to the nearly 0 percent error rate for machine-scored multiple choice tests. Such a high rate of error lowers confidence in the results and makes reporting of individual scores problematic, as most experts agree that reliabilities must have less than 10 percent measurement error to yield reliable individual results for diagnostic decision-making purposes (Salvia & Ysseldyke, 2001). In sum, objective scoring rubrics are essential to minimizing scorer subjectivity, thus leading to reliable and accurate scoring of essay exams.

How do test developers know what the national average is? How do I know how my child did in comparison to his classmates? Does it matter whether my child is compared to others his own age or in his own grade?

National norms are constructed by testing a representative sample of students from throughout the country. The sample is usually stratified in accordance with the most recent U.S. census to ensure that students in the sample are represented in proportion similar to their occurrence in the general U.S. population (Anastasi & Urbina, 1997). Samples are generally stratified to ensure representation based on sex, race, socioeconomic level (as determined by family income, parent education, or occupation), residence (urban, suburban, or rural), and geographic area of the country. The norm represents an average score for all students of a given age or grade level. Generally, it matters only slightly which

norm is used, an age norm or a grade norm; however, if the child is much older or younger than the average child in the grade, the differences in derived scores may have varying consequences. For example, a student who is very young for his or her grade, being less mature than the other students in the class, may not fare as well as the older students. These variations generally become less pronounced as students become older and abilities, rather than maturity, become more important.

On a norm-referenced test the derived scores will determine whether a comparison is being made among students with like characteristics. For example, if a percentile rank or stanine is reported, a comparison to age-mates or grade-mates is being made. If the test is criterion-referenced, as are many school performance tests, the comparison is made with a given standard of mastery (i.e., pass/fail, mastery/emerging/nonmastery), rather than age-mates or grade-mates (Thorndike, 1997).

Are these tests realistic measures of my child's knowledge in a particular subject area? How do these tests help identify my child's strengths and weaknesses? My child gets a single score or grade on all his other school tests, so why do they put bars on the student results graph to give a range of scores rather than a single score?

If students are motivated to perform to the best of their abilities, the test questions accurately measure the domain of knowledge, and testing conditions do not interfere with test performance, then the assessment most likely will accurately depict student performance in a given subject or ability area (Salvia & Ysseldyke, 2001). Most standardized tests provide a score for several subject areas, and this helps determine whether a student displays significant strengths or weaknesses in the areas assessed.

Tests measure strengths and weaknesses in two ways: interpersonal and intrapersonal. Interpersonal strengths and weaknesses are determined by comparing how a student performed compared to age- or grade-mates who took the same test. For interpersonal strengths and weaknesses, a cutoff score is determined and used for decision-making purposes. For example, students performing below the 25th percentile rank may be categorized "at risk" or in need of remedial services. Thus, any score at or below the 25th percentile rank would be considered an interpersonal weakness.

Intrapersonal strengths and weaknesses compare a child's

performance in one skill area on a test to the same child's performance in all other skill areas, to see where particular talents or difficulties lie. To determine intrapersonal weaknesses, an overall average is sometimes provided, or the test scores can be averaged manually. Percentile ranks cannot be averaged because they are not equal-interval units of measurement (Anastasi & Urbina, 1997). Percentile ranks must be converted to standardized scores to be averaged, then converted back to percentile ranks. Significant deviations (strengths if the deviations are above the mean, weaknesses if the deviations are below the mean) can then be determined. A significant deviation is often determined to be one standard deviation (or a given number of standard score points) above or below the average test performance.

The bars on a summary graph are derived from a statistical concept known as *standard error of measurement* (SEM). SEM is based on a test's reliability; the more reliable a test, the smaller the bar, the less reliable the test the larger the bar (Thorndike, 1997). If a test is perfectly reliable, the bar comprises the single score the student obtained on the test. SEM is essential when considering a student's score because, contrary to popular opinion, the score a student receives on a test is usually not the "true" score because no test is perfectly reliable. Thus, it is best to consider that a student's true score falls within a range of scores, as determined by the SEM, or the bar on the graph.

Furthermore, scores can be reported at different levels of confidence (Cohen & Swerdlik, 1999). For example, if a score is reported at a 68 percent level of confidence, then given 100 alternate-form administrations of the test, the student's true score likely falls within the given range 68 times. Thus, if the child's deviation IQ score was 93 and the SEM equals 5 standard score points, with a 68 percent level of confidence, the student's true IQ is likely to fall within the IQ range of 88 to 98 (or 93 ± 5) on 68 of 100 administrations of the IQ test. Although the 68 percent level of confidence is the range most commonly reported for scores, using this level of confidence means the student's score will fall *outside* the given range on about one of every three administrations—that is, the range will be wrong 32 percent of the time. Therefore, it is better to use two SEMs to report scores at the 95 percent level of confidence. Such a range (83–103, or 93 ± 10) means the student's true score will fall outside the given range only about one time in 20, giving far more confidence in the results and in the subsequent decisions.

What does it mean when scores are very different in different areas?

This is generally an indication that the student displays relative intrapersonal strengths and weaknesses. The weaknesses often require remediation, either through additional instruction, tutoring, remedial academic services, or special education services.

Does this test reflect my child's true performance or can outside conditions, like illness or anxiety, affect the test scores?

External conditions, such as noises and illness, as well as internal factors, such as anxiety or motivation, can affect test scores for certain children. On the other hand, many children are resilient and capable of maintaining focus and attention under conditions others would find distracting. Some testing conditions have been shown to adversely affect student performance, including poor lighting, insufficient workspace, uncomfortable seating, interruptions during timed tests, and the demeanor of the examiner (Anastasi & Urbina, 1997). Even the type of response format (e.g., marking an answer on the page versus coloring in an answer bubble) can affect scores for students in grades four or lower. Illness can certainly affect performance, although the effects are child-specific.

Anxiety is a different matter. The Yerkes-Dodson law (Schafer, 1996) indicates that moderate anxiety actually maximizes student performance. Low anxiety tends to result in low performance because it usually reflects low motivation. High anxiety often leads to low performance because the student feels overwhelmed. Indeed, test anxiety (test phobia, test fright) is a common, treatable condition that may lower student performance in up to 10 percent of the school-age population.

Using Test Results

How will these tests affect my child's instruction or the school's curriculum? How will test results determine the amount of assistance my child gets or the quality of the school?

Schools and school systems differ markedly in how they use standardized test results to change curriculum and instructional practices, or to make decisions about individual student placement or services. In general, what happens as a result of student scores depends

on the purpose of the test. If the purpose is to assess the effectiveness of the instruction or curriculum, the effects on individual students may be minimal, and the effects on the instructors or the curriculum may be substantial. If the curriculum is being assessed, student performance is an indicator of how closely aligned the school's curriculum and the test are with national, state, or local standards, as well as how effective the instruction is in implementing the curriculum. Because of the factor of alignment, it is important not to conclude immediately that low test scores are the result of poor teaching.

If an individual student scores well on tests, educators often use this information to provide a more challenging curriculum, such as through advanced, honors, or gifted programs. If the student performs poorly, educators often use this information to provide more academic support, such as through tutoring, remedial academic services, or special education services.

The Consequences of Testing

How much emphasis do tests like the SAT or ACT Assessment have on college admissions decisions? Can the scores determine what kind of college or job my child is prepared for?

Scores on tests such as the SAT or ACT Assessment are given different emphases by different universities. Although most institutions of higher learning require students to take an entrance exam, institutions are placing increasingly less emphasis on test scores. It is important to note that tests like the SAT were designed to predict college success, primarily during the freshman year, and they do this quite well. In general, the exclusive, competitive universities require high test scores, as well as high grade point averages, class ranks, and so on. Scores on college entrance exams have little to do with the kind of job the student may attain in the future.

In general, competitive colleges put more emphasis on test scores because test scores are objective, level the playing field, and predict college success. Thus, entrance exams act as an excellent method of screening students to move a pool of candidates to the next level. It is at this next level that letters of recommendation, extracurricular activities, and GPA become essential.

Releasing Test Results

Who sees my child's test results?

The Family Educational Rights and Privacy Act of 1974 ensures that parents and guardians have the opportunity to review, challenge, and correct their children's school records. The right to review test scores included in a child's cumulative or permanent school record is guaranteed by this act. To facilitate the dissemination of test information, school systems often send copies of test scores directly to parents and guardians as soon as the scores become available.

Besides parents and guardians, all persons with a legitimate educational interest in a particular child have access to test scores (Salvia & Ysseldyke, 2001). This may include all teachers and educational specialists who work with the child, school administrators, and other school officials. Parents have the right to request a list of all people who have access to their child's test information.

The Family Educational Rights and Privacy Act also aims to control dissemination of student information. Under the stipulations of the law, test information cannot be released without the parent or guardian's consent (or the child's consent if he or she is 18 years or older) to anyone other than those who have a legitimate educational interest in the child. For example, parent or guardian's consent must be given before test results can be released to social service agencies, law enforcement, or insurance companies. If a subpoena is issued, however, the parent or guardian's consent is not required for the release of test information.

Fair Testing

Are the tests my child takes fair?

Tests should give all students an equal opportunity to demonstrate their ability and knowledge (Childs, 1990). If a test seems not to provide equal opportunity, issues of bias must be considered. Discussions about test bias usually arise around issues of ethnicity, race, and gender. Tests are considered to be biased if individuals of the same ability but different demographic characteristics obtain different scores. Test bias is a complex issue. It may be attributed to representation or lack of representation of diverse populations in assessment materials; test administration procedures; students' knowledge of the nature of

assessment; wording of test items; linguistic backgrounds; test format; or even stereotypes, attitudes, and values (Childs, 1990; Coffman & Lindquist, 1980; McMillan, 2000; Salvia & Ysseldyke, 2001).

Developers work hard to eliminate bias in tests, but no test is perfect. Although true bias must be uncovered through statistical analysis, parents and guardians can work with educators to monitor and reduce potential test bias by ensuring that all teachers and students involved in the testing understand and follow the test administration procedures; by eliminating test items or material that may be offensive to individuals of a particular ethnicity, race, or gender; and by eliminating references in a test to things or ideas that may be unfamiliar to individuals of a particular race, ethnicity, or gender.

Although issues of test bias are extremely important and complex, a more important issue may be the fair use of tests. Unfortunately, even unbiased tests can be used in unfair ways that either help or hinder particular groups of students (Childs, 1990). Perhaps the best way for educators, parents, and guardians to address the issue of fairness in testing is to ensure that a variety of tests are wisely used as components of a multifaceted assessment program (ERIC Clearinghouse on Urban Education, 2001; Garcia, 1986).

Special Accommodations

How are tests modified to meet the needs of students with special needs or different learning styles?

The primary purpose of testing is to benefit students (Salvia & Ysseldyke, 2001). To fulfill this aim, tests must be accessible and appropriate for all students. Public Law 94–142 directs schools and school systems to ensure that when a test is given to a child with a disability, the test results reflect the skills the test is supposed to measure, not the child's disability. If a test is designed to measure reading comprehension, for example, it should measure a child's ability to understand what he or she reads, not whether the child may have a visual impairment. Of course, if a test is designed to measure a child's disability, it should in fact do that.

The legal call for accurate and accessible assessment brings to light the need for test accommodations. A test accommodation "involves adapting or modifying measures to enable students with disabilities to participate in assessment" (Salvia & Ysseldyke, 2001, p. 180). Throughout the country, there is great variation in the kinds of test

accommodations made. Some frequently used test accommodations include modification in the test format (e.g., large print edition of the test, Braille edition of the test), modification in the response format (e.g., respond orally, respond using a computer), modification in the ways in which the test can be taken (e.g., in a small setting, alone), and modification in the timing of the test (e.g., extended time, over several sessions). Test accommodations may also include the use of technology. Recent court decisions in some states allow students with learning disabilities to use electronic spell checking and dictation machines on tests (Ediger, 2001).

A list of appropriate test accommodations could be endless. As research on learning styles and disabilities continues to grow and educational technology continues to advance, more specific questions about legitimate test accommodations will arise. What is more important than a list of acceptable accommodations, therefore, may be an understanding of the purpose of testing and the specific needs of an individual. When facing difficult questions about test accommodations, parents, guardians, and educators may best serve students by holding fast to the spirit of laws such as Public Law 94–142, which ensure appropriate and accessible education for all students.

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Bagin, C. B. (1989). Talking to your high school students about standardized tests. *Practical Assessment, Research & Evaluation, 1*(4). Retrieved from http://ericae.net/pare/getvn.asp?v=1&n=4.

Bond, L. A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation, 5*(2). Retrieved from http://ericae.net/pare/getvn.asp?v=5&n=2.

Childs, R. A. (1990). Gender bias and fairness. *Practical Assessment, Research & Evaluation, 2*(3). Retrieved from http://ericae.net/pare/getvn.asp?v=2&n=3.

Coffman, W. E., & Lindquist, E. F. (1980). The testing of educational achievement in children. *Journal of Negro Education, 49*(3), 312–325.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurement* (4th ed.). Mountain View, CA: Mayfield Publishing.

Ediger, M. (2001). *Taking tests: More time for the handicapped?* (ERIC Document Reproduction Service No. ED 455 276)

Educational Testing Service. (1999). *What every parent should know about testing.* Retrieved from http://www.ets.org/pta2.html.

Eissenberg, T. E., & Rudner, L. M. (1988). Explaining test results to parents. *Practical Assessment, Research & Evaluation, 1*(1). Retrieved from http://ericae.net/pare/getvn.asp?v=1&n=1.

Erford, B. T., Vitali, G., Haas, R., & Boykin, R. R. (1995). *Essential Skill Screener (ESS): At-risk identification.* East Aurora, NY: Slosson.

ERIC Clearinghouse on Urban Education. (2001). *Gender differences in educational achievement within racial and ethnic groups* (Report No. EDO-UD-01–3). Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED 455 341)

Garcia, P. A. (1986). The impact of national testing on ethnic minorities: With proposed solutions. *Journal of Negro Education, 55*(3), 347–357.

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications and issues* (5th ed.). Belmont, CA: Wadsworth.

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation, 7*(18). Retrieved from http://ericae.net/pare/getvn.asp?v=7&n=18.

Mehrens, W. A. (1989). Preparing students to take standardized achievement tests. *Practical Assessment, Research & Evaluation, 1*(11). Retrieved from http://ericae.net/pare/getvn.asp?v=1&n=11.

Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders* (3rd ed.). Needham, MA: Pearson Educational Company.

Salvia, J., & Ysseldyke, J. (2001). *Assessment* (8th ed.). Boston, MA: Houghton Mifflin.

Schafer, W. (1996). *Stress management for wellness* (3rd ed.). Fort Worth, TX: Harcourt Brace.

Thorndike, R. L. (1997). *Measurement and evaluation in psychology and education* (6th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").